Information bounds and efficient estimation in a class of censored transformation models

Dorota M. Dabrowska
Department of Biostatistics
University of California
Los Angeles, CA 90095-1772
dorota@ucla.edu

February 2, 2008

Abstract

Transformation models provide a popular tool for regression analysis of censored failure time data. The most common approach towards parameter estimation in these models is based on nonparametric profile likelihood method. Several authors proposed also ad hoc M-estimators of the Euclidean component of the model. These estimators are usually simpler to implement and many of them have good practical performance. In this paper we consider the form of the information bound for estimation of the Euclidean parameter of the model and propose a modification of the inefficient M-estimators to one-step maximum likelihood estimates.

1 Introduction

The proportional hazard model, originating in Lehmann [17] and Cox [8], provides the most common tool for regression analysis of failure time data. The simplest version of this model assumes that a failure time T and a covariate Z satisfy the interrelationship

$$h(T) = -\beta^T Z + \varepsilon \tag{1.1}$$

where h is an unknown continuous increasing function mapping the support of T onto the real line, β is an unknown regression coefficient and ε is an error term independent of the covariate Z and having extreme value distribution with density $f(x) = e^x \exp[-e^x]$. Alternatively, if μ is the marginal distribution of the covariate and $\overline{A}(t,z)$ the cumulative hazard function of the conditional distribution of the failure time T given $Z = z$, then the model stipulates that

$$\overline{A}(t,z) = \Gamma(t)e^{\beta^T z} \quad \mu \text{ a.e. } z, \tag{1.2}$$

where β is an unknown regression coefficient and Γ is an unknown continuous increasing function, $\Gamma(0) = 0$, mapping the support of the failure time T onto the positive half line.

During the past two decades several authors proposed generalizations of (1.1)-(1.2) to semiparametric transformation models specifying the interrelationship between the conditional hazard function $\overline{A}(t,z)$ and the transformation $\Gamma(t)$ as

$$\overline{A}(t,z) = A(\Gamma(t), \theta|z) \quad \mu \text{ a.e. } z, \tag{1.3}$$

where $\{A(\cdot, \theta|z) : \theta \in \Theta\}$ is a family of conditional cumulative hazards dependent on a finite dimensional parameter θ. Following Bickel *et al.* [4], this family is referred to as the "core model". Common choices of (1.3) include scale regression models with core model derived from distributions with decreasing hazard rates, such as the frailty distributions with finite mean. In particular, the proportional odds ratio model has gained much popularity as a competitor to the proportional hazard model [1, 9, 10, 15, 18, 21, 23]. Core models with increasing or non-monotone hazards were considered in [5, 6, 7, 9, 12].

In this paper we consider estimation of the parameter θ based on an iid sample of right-censored failure times. For purposes of analysis of the odds ratio model, Murphy *et al.* [18] proposed to use nonparametric profile likelihood method. The approach taken was similar to the classical proportional hazard model. The model (1.3) was extended to include all monotone functions. With fixed parameter θ, an approximate likelihood function for the pair (θ, Γ) was maximized with respect to Γ to obtain an

estimate $\Gamma_{n\theta}$ of the unknown transformation. The estimate $\Gamma_{n\theta}$ was shown to form a step function placing mass at each uncensored observation, and the parameter θ was estimated by maximizing the resulting profile likelihood. Under certain regularity conditions on the censoring distribution, the authors showed that the estimates are consistent, asymptotically Gaussian at rate $\sqrt{n}$. They also proposed to estimate the standard errors of the regression coefficients by numerically twice-differentiating the log-profile likelihood. The approach was generalized to other transformation models whose core models have decreasing hazards in [16, 21, 22].

In this paper we take a different approach towards construction of efficient estimators of the parameter θ. To motivate it, let us recall [4] that in a regular parametric model $\mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{H}\}$, the asymptotic variance of any regular estimator of the parameter θ satisfies the bound

$$\mathrm{var}[\sqrt{n}(\widehat{\theta} - \theta)] \geq [I_{11}(\theta,\eta) - I_{12}(\theta,\eta)I_{22}(\theta,\eta)^{-1}I_{21}(\theta,\eta)]^{-1},$$

where

$$I(\theta,\eta) = \begin{pmatrix} I_{11}(\theta,\eta), I_{12}(\theta,\eta) \\ I_{21}(\theta,\eta), I_{22}(\theta,\eta) \end{pmatrix}$$

is the Fisher information matrix with entries

$$\begin{aligned} I_{11}(\theta,\eta) &= E\dot{\ell}_\theta(X,\theta,\eta)^{\otimes 2} \quad I_{22}(\theta,\eta) = E\dot{\ell}_\eta(X,\theta,\eta)^{\otimes 2}, \\ I_{12}(\theta,\eta) &= E\dot{\ell}_\theta(X,\theta,\eta)\dot{\ell}_\eta(X,\theta,\eta)^T = I_{21}(\theta,\eta)^T. \end{aligned}$$

Here $\dot{\ell}_\theta$ and $\dot{\ell}_\eta$ represent score functions corresponding to the two parameters. Alternatively,

$$\mathrm{var}[\sqrt{n}(\widehat{\theta} - \theta)] \geq \left[E\ell^*(X,\theta,\eta)\ell^*(X,\theta,\eta)^T\right]^{-1},$$

where ℓ^* is the efficient score function for estimation of θ. The function ℓ^* is the (componentwise) projection of the vector of scores $\dot{\ell}_\theta$ onto the orthocomplement of the nuisance tangent space $\dot{\mathcal{P}}_\eta$ spanned by all scores of the nuisance parameter η. To estimate the parameter θ, we may consider solving the score equation

$$\frac{1}{n}\sum_{i=1}^{n} \ell^*(X_i,\theta,\widehat{\eta}(\theta)) = o_P(n^{-1/2}), \tag{1.4}$$

where $\widehat{\eta}(\theta)$ is an estimate of η obtained "for each fixed θ" in the parameter set Θ. To be more precise, we assume that $\mathcal{P}$ is a submodel of a larger family of distributions $\mathcal{Q}$, and there exists a parameter $\widetilde{\eta} : \mathcal{Q} \times \Theta \to \mathcal{H}$, with the property $\widetilde{\eta}(Q,\theta) = \eta$, whenever $Q = P_{\theta,\eta} \in \mathcal{P}$. We require the estimate $\widehat{\eta}(\theta)$ to be asymptotically unbiased for estimation of $\widetilde{\eta}(Q,\theta)$ in the larger model $\mathcal{Q}$. If the equation (1.4) has a consistent solution $\widehat{\theta}$, then $\widehat{\theta}$ is an efficient estimate of the parameter θ under additional conditions

on $\widehat{\eta}(\widehat{\theta})$) given in [4, Ch.7.7]. These amount to the assumption that $\ell^*(\cdot, \theta, \widehat{\eta}(\widehat{\theta}))$ is a consistent estimator of the efficient score function and the bias of this estimator converges in probability to 0 at a rate faster than $\sqrt{n}$.

Turning to transformation models, in the case of uncensored data Bickel [2], Bickel and Ritov [3] and Klaassen [15] used invariance of the model with respect to the group of increasing transformations to show that the efficient score function for estimation of θ is given by a nonlinear rank statistics. Its form was derived using Sturm-Liouville theory. In the case of censored data, several authors verified existence of $\sqrt{n}$ estimators of the unknown transformation and used them to construct ad hoc estimators of the parameter θ. Whereas these estimators are inefficient, many of them have good practical performance [5, 6, 7, 23] and are simpler to implement than the profile likelihood method. They also apply to a wider class of transformation models.

In section 2, we assume the so-called "non-informative censoring model" and consider core models whose hazard rates are supported on the whole positive half-line, and are finite and positive at $x = 0$. Under a certain integrability condition, we derive the form of the information bound and efficient score function for estimation of the parameter θ. In section 3 we verify the integrability condition in the special case of the generalized odds ratio and the linear hazard regression models. In Section 4 we construct a class of Z estimators of the parameter θ. To study its properties, we shall make the assumption that the censoring distribution has support contained in the support of the failure time distribution and its upper point forms an atom. Under mild regularity conditions, we show that the proposed Z-estimators have an asymptotic distribution not depending on the choice of the estimator of the unknown transformation. As a by-product, we also show that the parameter θ can be efficiently estimated by solving an equation of the form (1.1) or by means of one-step MLE.

2 Information bound

2.1 Martingale identities

Throughout the paper we assume that the triple (X, δ, Z) represents a nonnegative withdrawal time (X), a binary withdrawal indicator ($\delta = 1$ for failure and $\delta = 0$ for loss-to-follow-up) and a vector of covariates (Z). The triple (X, δ, Z) is defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and $X = T \wedge T'$ and $\delta = 1(X = T)$, where T and T' represent failure and censoring times, respectively. We assume that T and T' are

conditionally independent given Z.

We denote by

$$\begin{aligned}\overline{A}(t,Z) &= \int_0^t P(X\in du,\delta=1|X\geq u,Z),\\ \overline{A}_c(t,Z) &= \int_0^t P(X\in du,\delta=0|X\geq u,Z),\end{aligned}$$

the conditional cumulative hazard functions of the failure and censoring time and set

$$\begin{aligned}N(t) &= 1(X\leq t,\delta=1),\quad N_c(t)=1(X\leq t,\delta=0),\\ \Lambda(t) &= \int_0^t Y(u)\overline{A}(du,Z),\quad \Lambda_c(t)=\int_0^t Y(u)\overline{A}_c(du,Z),\end{aligned}$$

where $Y(u)=1(X\geq u)$. Then $M(t)=N(t)-\Lambda(t)$ and $M_c(t)=N(t)-\Lambda_c(t)$ form mean zero martingales with respect to the self-exciting filtration $\{\mathcal{F}_t : t\geq 0\}$, $\mathcal{F}_t=\sigma\{1(X\leq t)\delta,1(X\leq t)X,1(X\leq t)(1-\delta),Z\}$.

Let $\tau_0=\sup\{t : EY(t)>0\}$ and let Q and Q_c denote the joint subdistribution functions

$$Q(t,z)=P(X\leq t,\delta=1,Z\leq z)\quad\text{and}\quad Q_c(t,z)=P(X\leq t,\delta=0,Z\leq z).$$

Then the failure and censoring counting processes satisfy

$$\begin{aligned}&\int_0^{\tau_0} h(u,Z)M(du)\in L_2^0(P)\quad\text{iff}\quad h\in L_2(Q),\\ &\int_0^{\tau_0} h(u,Z)M_c(du)\in L_2^0(P)\quad\text{iff}\quad h\in L_2(Q_c).\end{aligned}$$

Assuming that the function $\overline{A}$ is continuous, the two martingale processes are orthogonal. Nan, Edmond and Wellner [19] showed also that any function $b\in L_2^0(P)$ can be represented as a sum

$$b(X,\delta,Z)=\int_0^{\tau_0} R_1[b](u,Z)M(du)+\int_0^{\tau_0} R_2[b](u,Z)M_c(du)+E[b|Z],$$

where $R_1[b]\in L_2(Q)$ and $R_2[b]\in L_2(Q_c)$ are given by

$$\begin{aligned}R_1[b](x,z) &= b(x,1,z)-E[b(X,\delta,Z)|X\geq x,Z=z],\\ R_2[b](x,z) &= b(x,0,z)-E[b(X,\delta,Z)|X\geq x,Z=z].\end{aligned}$$

The three terms in this representation are orthogonal.

2.2 Assumptions and notation

To derive the form of the information bound and efficient score function in the transformation model, we make the following regularity conditions on the core model.

Condition 2.1 (i) The parameter set $\Theta \subset R^d$ is open, and the parameter θ is identifiable in the core model: $\theta \neq \theta'$ iff $A(\cdot, \theta, z) \not\equiv A(\cdot, \theta', z)$ for μ–a.e. z.

(ii) For μ almost all z, the function $A(\cdot, \theta, z)$ has a hazard rate $\alpha(\cdot, \theta, z)$ supported on the whole positive half-line and there exist constants $0 < m_1 < m_2 < \infty$ such that $m_1 \le \alpha(0, \theta, z) \le m_2$ for μ–a.e. z and all $\theta \in \Theta$.

(iii) The function $\ell(x, \theta, z) = \log \alpha(x, \theta, z)$ is continuously differentiable with respect to both x and θ.

(iv) The family $\{\alpha(x, \theta, z) : \theta \in \Theta\}$ forms a regular parametric model, in particular information is finite and positive definite.

The derivatives of $\ell(x, \theta, Z)$ are denoted by

$$\dot\ell(x, \theta, Z) = \frac{\partial}{\partial \theta}\ell(x, \theta, Z) \quad \text{and} \quad \ell'(x, \theta, Z) = \frac{d}{dx}\ell(x, \theta, Z).$$

Under assumption of transformation model, the true distribution P of (X, δ, Z) belongs to the family $\mathcal{P} = \{P_{(\theta,\eta)} : \theta \in \Theta, \eta \in \mathcal{H}\}$ where θ is the Euclidean parameter of interest and $\eta = (\Gamma, \overline{A}_c, \mu)$ is the nuisance parameter corresponding to the unknown transformation Γ, the unknown cumulative hazard function $\overline{A}_c(t; z)$ of the conditional distribution of the censoring time given the covariate Z and the marginal distribution μ of Z. We make the following regularity conditions.

Condition 2.2 Let $\tau_0 = \sup\{t : P(X \ge t) > 0\}$ and $\tau_F = \sup\{t : P(T \ge t) > 0\}$.

(i) The distribution μ is nondegenerate.

(ii) The parameters $\overline{A}_c$ and μ are noninformative on (θ, Γ).

(iii) Γ is an increasing continuous function, $\Gamma(0) = 0$ and $\lim_{t \uparrow \tau_F} \Gamma(t) = \infty$.

(iv) If $\tau_0 < \tau_F$ then τ_0 is an atom of the marginal survival function of the censoring time.

The condition (iii) refers to the unobserved model corresponding to uncensored data. If $\tau_0 < \tau_F$, then the condition (iv) serves to ensure that the the transformation can be estimated consistently only within the range $[0, \tau_0]$. We do not know whether there exists a consistent estimate of the parameter θ if $\tau_0, \tau_0 < \tau_F$, is a continuity point of the censoring distribution.

Using conditions 2.1-2.2 and the assumption of the conditional indpendence of failure and censoring times, the score operators for the four parameters are

$$\begin{aligned}
\dot{\ell}_1[\theta](X,\delta,Z) &= \int_0^{\tau_0} \dot{\ell}(\Gamma(u),\theta,Z)M(du),\\
\dot{\ell}_2[g](X,\delta,Z) &= \int_0^{\tau_0} \left[g(u) + \ell'(\Gamma(u),\theta,Z)\left(\int_0^u g d\Gamma\right)\right] M(du),\\
\dot{\ell}_3[b](X,\delta,Z) &= \int_0^{\tau_0} b(u,Z)dM_c(u),\\
\dot{\ell}_4[c](X,\delta,Z)) &= c(Z),
\end{aligned}$$

where $c(Z) \in L_2^0(\mu)$, $b(u,Z) \in L_2(Q_c)$ and

$$g(X) + \ell'(\Gamma(X),\theta,Z)\int_0^X g d\Gamma \in L_2(Q). \tag{2.1}$$

The tangent spaces for the four parameters are $\dot{\mathcal{P}}_i = [\dot{\ell}_i], i = 1,\ldots,4$, where $[\alpha]$ denotes the closed linear span of the set α in $L_2^0(P)$. The spaces $\dot{\mathcal{P}}_i, i = 2,3,4$ are mutually orthogonal, and so are the spaces $\dot{\mathcal{P}}_i, i = 1$, and $i = 3,4$. The nuisance tangent space is $\dot{\mathcal{P}}_\eta = \dot{\mathcal{P}}_2 + \dot{\mathcal{P}}_3 + \dot{\mathcal{P}}_4$.

In the case of the proportional hazard model,

$$\overline{A}(t,z) = \Gamma(t)e^{\theta^T Z}, \tag{2.2}$$

Sasieni [20] and Nan, Edmond and Wellner [19] showed that the martingale oparator

$$U(f) = \int_0^{\tau_0} \left(f(x,Z) - E[f(X,Z)|X = x, \delta = 1]\right) M(dx), \quad f \in L_2(Q),$$

satisfies (i) $U(f) \perp \dot{\mathcal{P}}_\eta$ in $L_2(P)$; (ii) $\Pi(f|\dot{\mathcal{P}}_\eta^\perp) = U(R_1(f))$ for $f \in L_2^0(P)$ and (iii) $\dot{\mathcal{P}}_\eta^\perp = \{U(f) : f \in L_2(Q)\}$. Here $\Pi(\cdot|\dot{\mathcal{P}}_\eta^\perp)$ denotes the projection onto $\dot{\mathcal{P}}_\eta^\perp$, the ortho-complement of the nuisance tangent space. These properties entail that the efficient

score function for estimation of the parameter θ in the model (2.2) is given by $U(f)$, $f(x, Z) = Z$.

To obtain an extension of this result to the transformation model, we shall use the following notation. Firstly, we shall find it convenient to denote the "true" parameters (θ, Γ) as (θ_0, Γ_0). Further, let

$$\begin{aligned} s[1](u, \Gamma, \theta) &= EY(u)\alpha(\Gamma(u), \theta, Z) \\ s[f](u, \Gamma, \theta) &= EY(u)f(u, Z)\alpha(\Gamma(u), \theta, Z), \\ e[f](u, \Gamma, \theta) &= \frac{s[f]}{s[1]}(u, \Gamma, \theta) \end{aligned}$$

and

$$\mathrm{cov}[f_1, f_2] = e[f_1 f_2^T] - e[f_1]e[f_2]^T\ , \quad \mathrm{var}[f] = \mathrm{cov}[f, f]$$

Note that if failure and censoring times are conditionally independent given Z, then under the assumption of the transformation model, the conditional distribution of Z given $X = t$ and $\delta = 1$ has form

$$\begin{aligned} P(Z \in B | X = t, \delta = 1) &= \frac{1}{s[1](t, \Gamma_0, , \theta_0)} \int_B G_c(t, z)F(\Gamma_0(t), \theta_0, z)\alpha(\Gamma_0(t), \theta_0, z)\mu(dz) \\ &= e[f](t, \Gamma_0, \theta_0), \end{aligned}$$

where $f(t, Z) = 1(Z \in B)$, $F(x, \theta_0, z) = [\exp -A(x, \theta_0, z)]$ is the survival function of the core model and G_c is the conditional survival function of the censoring time. More generally,

$$e[f](t, \Gamma_0, \theta_0) = E[f(X, Z)|X = t, \delta = 1]$$

and similarly, $\mathrm{var}[f](t, \Gamma_0, \theta_0)$ and $\mathrm{cov}_0[f_1, f_2](t, \Gamma_0, \theta_0)$ are conditional variance and covariance operators. From section 2.1, we also have

$$f(X, Z) \in L_2(Q) \quad \text{iff} \quad E\int Y(u)f^2(u, Z)\alpha(\Gamma_0(u), \theta_0, Z)\Gamma_0(du) < \infty.$$

Using domintated convergence theorem,

$$f(X, Z) \in L_2(Q) \quad \text{iff} \quad s[f^2](u, \Gamma_0, \theta_0) \in L_1(\Gamma_0).$$

By noting that

$$\Gamma_0(t) = \int_0^t \frac{EN(du)}{s[1](u, \Gamma_0, \theta_0)},$$

the square integrability condition reduces to

$$f(X, Z) \in L_2(Q) \quad \text{iff} \quad e[f^2](u, \Gamma_0, \theta_0) \in L_1(Q).$$

Next, with some abuse of notation, we shall write $s[\ell'](u,\Gamma,\theta), e[\ell'](u,\Gamma,\theta)$ and $\mathrm{var}[\ell'](u,\Gamma,\theta)$ whenever $f(X,Z) = \ell'(\Gamma(X),\theta,Z)$. Let

$$\begin{aligned}
C(t) &= \int_0^t (s[1](u,\Gamma,\theta))^{-2}\, EN(du), \\
B(t) &= \int_0^t \mathrm{var}[\ell'](u,\Gamma,\theta) EN(du), \\
\mathcal{P}(u,t) &= \exp - \int_u^t s[\ell'](u,\Gamma,\theta) C(du), \qquad (2.3)\\
K(t,t') &= \int_0^{t\wedge t'} C(du)\mathcal{P}(u,t)\mathcal{P}(u,t')\ .
\end{aligned}$$

Finally, let $D[f], D[f](t) = D[f](t,\Gamma,\theta)$, denote the solution to the linear Volterra equation

$$D[f](t) = -\int_0^t s[f](u,\Gamma,\theta)C(du) - \int_0^t D[f](u-)s[\ell'](u,\Gamma,\theta)C(du)$$

The equation has a unique locally bounded solution given by

$$D[f](t) = -\int_0^t s[f](u,\Gamma,\theta)C(du)\mathcal{P}(u,t).$$

We shall need the following integrability condition.

Condition 2.3 Let

$$\kappa(\tau_0) = \int\int_{0<u\le t\le \tau_0} C(du)\mathcal{P}(u,t)^2 B(du)$$

and suppose that $\kappa(\tau_0) < \infty$.

Lemma 2.1 If the condition 2.3 holds then $K \in L_2(B\otimes B)$ and $D[f] \in L_2(B)$, for any function $f(X,Z)$ such that $e[f^2] \in L_1(Q)$

This lemma can be verified using Cauchy-Schwartz inequality and dominated convergence theorem. If failure and censoring times are conditionally independent and the transformation model is satisfied by $(\theta,\Gamma) = (\theta_0,\Gamma_0)$, then the second part of the lemma holds for any $f(X,Z) \in L_2(Q)$.

2.3 Information bound

Unless this leads to confusion, we suppress the dependence of the functions $s[f](\cdot,\Gamma,\theta)$, $e[f](\cdot,\Gamma,\theta)$ and the corresponding variance and covariance operators on (Γ,θ). We write $s_0[f], e_0[f], \mathrm{var}_0[f]$ and $\mathrm{cov}_0[f_1,f_2]$ whenever the failure and censoring times are conditionally independent given Z and the transformation model is satisfied by $(\Gamma,\theta)=(\Gamma_0,\theta_0)$. The functions $C_0, B_0, K_0, \mathcal{P}_0$ and D_0 are defined analogously.

Further, for any $\varphi \in L_2(B)$ and $f(X,Z)$ such that $e[f^2]\in L_1(Q)$, set

$$\rho[f,\varphi](u) = \mathrm{cov}[f,\ell'](u) - \varphi(u)\mathrm{var}[\ell'](u) \tag{2.4}$$

and let

$$\begin{aligned} W_f(t,\theta,Z) &= f(t,Z) - e[f](t) - (\ell'(\Gamma(t),\theta,Z) - e[\ell'](t))\,\varphi^*(t) \\ &- \frac{1}{s[1](t)}\int_t^{\tau_0} \mathcal{P}(t,s)\rho[f,\varphi^*](s)EN(ds), \end{aligned} \tag{2.5}$$

where φ^* is the solution to the Fredholm equation

$$\varphi^*(t) = -D[f](t) - \int_0^{\tau_0} K(t,u)\varphi^*(u)B(du) + \int_0^{\tau_0} K(t,u)\mathrm{cov}[f,\ell'](u)EN(du). \tag{2.6}$$

If $\mathrm{var}[\ell']\equiv 0$, then $W_f(t,\theta,Z) = f(t,Z) - e[f](t)$. In addition, setting $\psi^* = \varphi^* + D[f]$, the equation (2.6) simplifies to

$$\psi(t) - \lambda\int_0^{\tau_0} K(t,u)\psi(u)B(du) = \eta(t), \tag{2.7}$$

where $\lambda = -1$ and

$$\begin{aligned} \eta(t) &= \int_0^{\tau_0} K(t,u)\rho[f,-D[f]](u)EN(du), \\ \rho[f,-D[f]](u) &= \mathrm{cov}[f,\ell'](u) + \mathrm{var}[\ell'](u)D[f](u) \end{aligned}$$

If $\mathrm{var}[\ell']\not\equiv 0$ then the kernel K is symmetric, positive definite and square integrable with respect to B. Therefore it can have only positive eigenvalues. For $\lambda=-1$, the equation (2.7) has a unique solution given by

$$\psi^*(t) = \eta(u) - \int_0^{\tau_0} \Delta(t,u,-1)\eta(u)B(du), \tag{2.8}$$

where $\Delta(t,u,\lambda)$ is the resolvent corresponding to the kernel K. The resolvent equations

$$\begin{aligned} K(t,u) &= \Delta(t,u,\lambda) - \lambda \int_0^{\tau_0} \Delta(t,w,\lambda) B(dw) K(w,u) \\ &= \Delta(t,u,\lambda) - \lambda \int_0^{\tau_0} K(t,w) B(dw) \Delta(w,u,\lambda), \end{aligned}$$

applied with $\lambda = -1$ imply

$$\psi^*(t) = \varphi^*(t) + D[f](t) = \int_0^{\tau_0} \Delta(t,u,-1) \rho[f,-D[f]](u) EN(du). \tag{2.9}$$

If $\mathrm{var}[\ell'] \not\equiv 0$ but $\rho[f,-D[f]] \equiv 0$, then the solution to this equation is $\psi^* \equiv 0$, or equivalently, $\varphi^* = -D[f]$.

Define

$$U(f,\theta) = \int_0^{\tau_0} W(f,\theta,Z)(u) M(du)$$

Proposition 2.1 Suppose that the conditions 2.1-2.3 are satisfied with $(\theta,\Gamma) = (\theta_0,\Gamma_0)$. Then (i) $U(f,\theta_0) \perp \dot{\mathcal{P}}_\eta$ in $L_2(P)$; (ii) $\Pi(f|\dot{\mathcal{P}}_\eta^\perp) = U(R_1(f),\theta_0)$ for $f \in L_2^0(P)$ and (iii) $\dot{\mathcal{P}}_\eta^\perp = \{U(f,\theta_0) : f \in L_2(Q)\}$.

Proof . Set

$$\begin{aligned} g^*(t) &= e_0[f](t) - e_0[\ell'](t)\varphi^*(t) \\ &- \frac{1}{s_0[1](t)} \int_t^{\tau_0} \mathcal{P}_0(t,s) \rho_0[f,\varphi^*](s) EN(ds). \end{aligned}$$

Then

$$\varphi^*(t) = \int_0^t g^* d\Gamma_0$$

and

$$U(f,\theta_0) = \int_0^{\tau_0} f(t,Z) M(du) - \dot{\ell}_2[g^*] = \int_0^{\tau_0} [f(t,Z) - g^*(t) - \ell'(\Gamma_0(u),\theta_0,Z)\varphi^*(u)] M(du).$$

Using Cauchy-Schwartz inequality and the condition 2.3, it is easy to verify that $\dot{\ell}_2[g^*] \in L_2^0(P)$ for any function $f(X,Z) \in L_2(Q)$. Therefore part (i) of the proposition will be verified if we show that $E(U(f,\theta_0)\dot{\ell}_2[g]) = 0$ for any function g satisfying th condition (2.1).

Put $\dot{\ell}_2[g] = I_1(g) + I_2(g)$, where

$$\begin{aligned}
I_1(g) &= \int_0^{\tau_0} g(u)dM(u), \\
I_2(g) &= \int_0^{\tau_0} \left[\ell'(\Gamma_0(u), \theta_0, Z)\int_0^u g d\Gamma_0\right] M(du).
\end{aligned}$$

Then

$$\begin{aligned}
EU(f,\theta_0)I_1(g) &= -\int_0^{\tau_0} g(t)\left[\int_t^{\tau_0} \mathcal{P}_0(t,s)\rho_0[f,\varphi^*](s)EN(ds)\right]\Gamma_0(t), \\
EU(f,\theta_0)I_2(g) &= \int_0^{\tau_0}\left[\rho_0[f,\varphi^*](t)\int_0^t g d\Gamma_0\right] dEN(t) \\
&- \int e_0[\ell'](t)\left(\int_t^{\tau_0}\mathcal{P}_0(t,s)\rho_0[f,\varphi^*](s)EN(ds)\right)\left(\int_0^t g d\Gamma_0\right) d\Gamma_0(t).
\end{aligned}$$

Using

$$\mathcal{P}_0(t,s) = 1 - \int_t^s s_0[\ell']dC_0(du)\mathcal{P}_0(u,s) = 1 - \int_t^s e_0[\ell'](u)d\Gamma_0(u)\mathcal{P}_0(u,s)$$

for $t \le s$, and applying Fubini theorem, we have

$$\begin{aligned}
EU(f,\theta_0)I_2(g) &= \int_0^{\tau_0} g(u)\left(\int_u^{\tau_0}\rho_0[f,\varphi^*](t)dEN(t)\right)d\Gamma_0(u) \\
&- \int_0^{\tau_0} g(u)\left(\int_u^{\tau_0} e_0[\ell'](t)\left(\int_t^{\tau_0}\mathcal{P}_0(t,s)\rho_0[f,\varphi^*](s)EN(ds)\right)d\Gamma_0(t)\right)d\Gamma_0(u) \\
&= \int_0^{\tau_0} g(u)\left(\int_u^{\tau_0}\rho_0[f,\varphi^*](t)dEN(t)\right)d\Gamma_0(u) \\
&- \int_0^{\tau_0} g(u)\left(\int_u^{\tau_0}\left(\int_t^s e_0[\ell'](t)d\Gamma_0(t)\mathcal{P}_0(t,s)\right)\rho_0[f,\varphi^*](s)EN(ds)\right)d\Gamma_0(u) \\
&= \int_0^{\tau_0} g(u)\left(\int_u^{\tau_0}\rho_0[f,\varphi^*](t)dEN(t)\right)d\Gamma_0(u) \\
&+ \int_0^{\tau_0} g(u)\left(\int_u^{\tau_0}[\mathcal{P}_0(u,s)-1]\rho_0[f,\varphi^*](s)EN(ds)\right)d\Gamma_0(u) = -EU(f,\theta_0)I_1(g).
\end{aligned}$$

This completes the proof of part (i). Part (ii) and part (iii) follows in the same way as in Nan, Edmond and Wellner [19]. $\square$

Since $U(f,\theta_0)$ is a martingale operator, we have

$$\Sigma_0(f,\theta_0) = EU(f,\theta_0)^2 =$$

$$= \quad E\int W_f^2(u,\theta_0,Z)Y(u)\alpha(\Gamma_0(u),\theta_0,Z)\Gamma_0(du).$$

The matrix $\Sigma_0(f,\theta_0)$ satisfies $\Sigma_0(f,\theta_0)=\Sigma_1(f,\theta_0)+\Sigma_2(f,\theta_0)$, where

$$\begin{aligned} \Sigma_1(f,\theta_0) &= \int_0^{\tau_0} \mathrm{var}_0[f-\ell'\varphi^*](u)EN(du), \\ \Sigma_2(f,\theta_0) &= \int_0^{\tau_0} C_0(du)\left[\int_u^{\tau_0} \mathcal{P}_0(u,t)\rho_0[f,\varphi^*](t)EN(du)\right]^{\otimes 2} . \end{aligned} \tag{2.10}$$

The conditional variance function is identically equal to 0, if and only if

$$f(t,Z)=\ell'(\Gamma_0(t),\theta_0,Z)\int_0^t hd\Gamma_0+a(t)\in L_2(Q) \tag{2.11}$$

and $a\equiv h$. For any function of the form (2.11), Fubini theorem yields

$$\begin{aligned} D_0[f](t) &= -\int_0^t a(u)d\Gamma_0(u)\mathcal{P}_0(u,t) \\ &- \int_0^t \left(\int_0^u hd\Gamma_0\right) e_0[\ell'](u)d\Gamma_0(u)\mathcal{P}_0(u,t) \\ &= -w(t)-\int_0^t hd\Gamma_0, \\ w(t) &= \int_0^t [a-h](u)d\Gamma_0(u)\mathcal{P}_0(u,t). \end{aligned}$$

Hence (2.6) reduces to

$$\int_0^t [g^*-h]d\Gamma_0+\int_0^{\tau_0} K_0(t,u)[\left[\int_0^u (g^*-h)d\Gamma_0\right] B_0(du)=w(t).$$

If $a\equiv h$ then the right-hand side of this equation is identically equal to 0, and correspondingly, $g^*=h$. Otherwise, we obtain

$$\varphi^*(t)=\int_0^t hd\Gamma_0+w(t)-\int_0^{\tau_0}\Delta(t,u,-1)w(u)B_0(du)$$

and

$$\mathrm{var}_0[f-\ell'\varphi^*](t)=\mathrm{var}_0[\ell'](t)[w(t)-\int_0^{\tau_0}\Delta(t,u,-1)w(u)B_0(du)]$$

If $\mathrm{var}_0[\ell']\not\equiv 0$, then the right-hand side is identically equal to 0 if and only if

$$w(t)=\int_0^{\tau_0}\Delta(t,u,-1)w(u)B(du).$$

In this case, the resolvent equations imply that $w\equiv 0$ so that $a\equiv h$.

3 Examples

In this section we verify the square integrability condition in two models. In particular, we show that the information bound applied to both censored and uncensored data.

Example 3.1 Generalized odds ratio model. The survival function of the core model is given by

$$\begin{aligned} F(x,\theta,Z) &= [1+\eta e^{\theta^T Z}x]^{-1-1/\eta} \quad \text{for} \quad \eta>0, \\ &= \exp[-e^{\theta^T Z}x] \qquad \text{for} \quad \eta = 0. \end{aligned}$$

The proportional hazard model corresponds to the choice $\eta = 0$ and the proportional odds model to $\eta = 1$. The hazard function of the core model given by

$$\alpha(x,\theta,Z) = e^{\theta^T Z}[1+\eta e^{\theta^T Z}x]^{-1}.$$

We have

$$\begin{aligned} \dot\ell(x,\theta,Z) &= Z[1+\eta e^{\theta^T Z}x]^{-1}, \\ \ell'(x,\theta,Z) &= -\eta e^{\theta^T Z}[1+\eta e^{\theta^T Z}x]^{-1}. \end{aligned}$$

If $|Z| \le d_0$ and $d_1 \le e^{\theta^T Z} \le d_2$, then

$$\begin{aligned} |\dot\ell(x,\theta,Z)| &\le d_0[1+\eta d_1 x]^{-1} \le d_0 \\ \eta d_1[1+\eta d_1 x]^{-1} &\le -\ell'(x,\theta,Z) \le \eta d_2[1+\eta d_2 x]^{-1} \end{aligned}$$

Note that $-\ell'(x,\theta,Z)$ is an increasing function of $\eta e^{\theta^T Z}$.

Suppose that (θ_0, Γ_0) is the "true" parameter of the transformation model. The preceding bounds imply

$$\mathcal{P}_0(u,t) \le \exp \int_u^t \frac{\eta d_2}{1+\eta d_2 \Gamma_0(v)} d\Gamma_0(v) = \frac{1+\eta d_2\Gamma_0(t)}{1+\eta d_2 \Gamma_0(u)}.$$

Next recall that if U is a random variable with finite variance and distribution function H, then

$$\mathrm{Var}(U) = \frac{1}{2}\int\int (w_1 - w_2)^2 H(dw_1)H(dw_2) \tag{3.1}$$

By noting that

$$\begin{aligned} & [\ell'(\Gamma_0(t),\theta_0,z_1)-\ell'(\Gamma_0(t),\theta_0,z_2)]^2 = \\ = \;& \frac{\eta^2[e^{\theta_0^T z_1}-e^{\theta_0^T z_2}]^2}{[1+\eta e^{\theta_0^T z_1}\Gamma_0(t)]^2[1+\eta e^{\theta_0^T z_2}\Gamma_0(t)]^2} \\ \leq \;& \eta^2[d_2-d_1]^2[1+\eta d_1\Gamma_0(t)]^{-4}, \end{aligned}$$

and using (3.1), we obtain $\mathrm{var}[\ell'](t,\Gamma_0,\theta_0) = O(1)[1+\eta d_1\Gamma_0(t)]^{-4}$. Since $[1+\eta d_2\Gamma_0(t)] \leq (d_2/d_1)[1+\eta d_1\Gamma_0(t)]$, we see that

$$\kappa(\tau_0) = O(1)\int\left[\int_0^t \frac{d\Gamma_0(v)}{s[1](v,\Gamma_0,\theta_0)}\frac{1}{[1+\eta d_2\Gamma_0(v)]^2}\right]\frac{s[1](t,\Gamma_0,\theta_0)d\Gamma_0(t)}{(1+\eta d_1\Gamma_0(t))^2}$$

In the generalized odds ratio model, $s[1](t,\Gamma_0,\theta_0)$ is a decreasing function of t. Therefore

$$\frac{s[1](t,\Gamma_0,\theta_0)}{s[1](u,\Gamma_0,\theta_0)} \leq 1$$

and

$$\kappa(\tau_0) \leq \widetilde{\kappa}(\infty) = O(1)\int[1-(1+\eta d_2 w))^{-1}](1+\eta d_1 w)^{-2}dw = O(1)$$

The bound is valid for any distribution of the censoring times. We obtain a better bound on the constant $\widetilde{\kappa}(\infty)$ in the case of the so-called Koziol-Green censoring model. The conditional survival function of the censoring time given the covariate is of the form

$$G_c(t|z) = F(\Gamma_0(t),\theta_0,z)^a, \quad a \geq 0.$$

The choice of $a = 0$ corresponds to the case of uncensored data. We have

$$s[1](t,\Gamma_0,\theta_0) = Ee^{\theta_0^T Z}[1+\eta e^{\theta_0^T Z}\Gamma_0(t)]^{-[1+(1+a)/\eta]}$$

so that

$$d_1[1+\eta d_2\Gamma_0(t)]^{-[1+(1+a)/\eta]} \leq s[1](t,\Gamma_0,\theta_0) \leq d_2[1+\eta d_1\Gamma_0(t)]^{-[1+(1+a)/\eta]}.$$

Hence

$$\begin{aligned} \widetilde{\kappa}(\infty) &= O(1)\int_0^\infty\left[\int_0^x (1+\eta d_2 w)^{(1+a)/\eta-1}dw\right](1+\eta d_1 x)^{-[3+(1+a)/\eta]}dx \\ &= O(1)\int_0^\infty (1+\eta d_1 x)^{-3}dx = O(1). \end{aligned}$$

Example 3.2 Linear hazard model has failure rate

$$\alpha(x,\theta|z) = a_\theta(z) + x b_\theta(z).$$

We assume that $m_1 \le a_\theta(z) \le m_2$, $m_1' \le b_\theta(x) \le m_2'$ for some finite positive constants $m_q, m_q', q = 1,2$. In addition, the functions $a_\theta(z)$ and $b_\theta(z)$ have bounded derivatives with respect to θ. We have

$$\ell'(x,\theta,z) = \frac{b_\theta(z)}{a_\theta(z) + x b_\theta(z)}$$

Set $d_1 = m_1'/m_2$ and $d_2 = m_2'/m_1$ and suppose that (θ_0, Γ_0) is the true parameter of the transformation model. Using a similar algebra as in the example 3.1, we can show that

$$\mathcal{P}_0(u,t) \le \frac{1 + d_1\Gamma_0(u)}{1 + d_1\Gamma_0(t)} \quad \mathrm{var}[\ell'](t,\Gamma_0,\theta_0) \le O(1)[1 + d_1\Gamma_0(t)]^{-4}.$$

Hence

$$\kappa(\tau_0) = O(1)\int_0^{\tau_0}\left[\int_0^t \frac{d\Gamma_0(v)}{s[1](v,\Gamma_0,\theta_0)}[1 + d_1\Gamma_0(v)]^2\right]\frac{s[1](t,\Gamma_0,\theta_0)d\Gamma_0(t)}{(1 + d_1\Gamma_0(t))^6}$$

For $v < t$, we have

$$\frac{s[1](t,\Gamma_0,\theta_0)}{s[1](v,\Gamma_0,\theta_0)} \le O(1)\frac{1 + d_2\Gamma_0(t)}{1 + d_1\Gamma_0(v)} = O(1)\frac{1 + d_1\Gamma_0(t)}{1 + d_1\Gamma_0(v)}$$

Hence

$$\kappa(\tau_0) \le \widetilde{\kappa}(\infty) = O(1)\int_0^\infty (1 + d_1 w)^{-3} dw$$

Other examples satisfying the integrability condition 2.3 include the inverse Gaussian core model, and half-symmetric distributions such as the half-normal, half-logistic and half-t distribution.

4 Estimation

We turn now to estimation of the parameter θ. Let us recall that under the assumption of transformation model, the true distribution of (X, δ, Z) is in a class $\mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{H}\}$, where η represents the triple $\eta = (\Gamma, G_c, \mu)$.

To construct an estimator of the parameter θ, we assume that $\mathcal{Q}$ is a class of probability distributions of the variables (X, δ, Z) containing $\mathcal{P}$ as a submodel. For each $(Q, \theta) \in \mathcal{Q} \times \Theta$, we let $\Gamma_{Q,\theta}$ be a monotone function such that

$$\Gamma_{Q,\theta_0} = \Gamma_0$$

if $Q = P_{\theta_0,\eta_0}$ and $\xi_0 = (\theta_0, \Gamma_0, G_c, \mu)$ is the true parameter of the transformation model.

Dropping dependence of this function on the distribution Q, let $\Gamma_{n,\theta}$ be an estimator of Γ_θ such that $\|\Gamma_{n\theta} - \Gamma_\theta\|_\infty = o_Q(1)$, i.e. the estimate is consistent when observations (X, δ, Z) are sampled from a distribution $Q \in \mathcal{Q}$. In addition to this we assume the following regularity conditions.

Condition 4.1 Let $B(\theta_0, \varepsilon_n)$ denote an open ball of radius ε_n and centered at θ_0.

(i) $\varepsilon_n \downarrow 0$ and $\sqrt{n}\varepsilon_n \uparrow \infty$.

(ii) The point $\tau_0 = \sup\{t : EY(t) > 0\}$ is an atom of the marginal distribution of the censoring times.

(iii) The estimate of the transformation satisfies: $\sqrt{n}\|\Gamma_{n0} - \Gamma_0\|_\infty = O_P(1)$, $\limsup_n\{\|\Gamma_{n\theta}\|_v : \theta \in B(\theta_0, \varepsilon_n)\} = O_P(1)$ and

$$\sup\{\sqrt{n}\|\Gamma_{n\theta} - \Gamma_{n0}\|_\infty / [\sqrt{n}|\theta - \theta_0| + 1] : \theta \neq \theta_0, \theta \in B(\theta_0, \varepsilon_n)\} = O_P(1)$$

Examples of estimators satisfying these conditions were given by Cuzick [9], Bogdanovicius and Nikulin [5] and Yang and Prentice [23], among others.

Referring to the notation of section 2, we assume that the function $f(u, Z)$ is of the form $f(u, Z) = f(\Gamma_\theta(u), \theta, Z)$ and make the following regularity conditions.

Condition 4.2 Let ψ be a constant or a bounded continuous strictly decreasing function. For $p = 1, 2, 3$, let ψ_p be continuous bounded or strictly increasing functions such that $\psi_p(0) < \infty$ and

$$\int_0^\infty e^{-x}\psi_1^2(x)dx < \infty, \quad \int_0^\infty e^{-x}\psi_2(x)dx < \infty, \quad \int_0^\infty e^{-x}\psi_3(x)dx < \infty\ .$$

Suppose that the derivatives of the function $\ell(x, \theta, z)$ satisfy

$$|\ell'(x, \theta, z)| \le \psi(x), \quad |\ell''(x, \theta, z)| \le \psi(x), \quad |\dot{\ell}(x, \theta, Z) \le \psi_1(x)$$

The function $f(x, \theta, Z)$ is differentiable with respect to x and

$$|f(x, \theta, Z)| \leq \psi_1(x), \qquad |f'(x, \theta, Z)| \leq \psi_2(x).$$

We also have

$$\begin{aligned}
&|g_1(x, \theta, z) - g_1(x', \theta, z)| \leq max[\psi_3(x), \psi_3(x')]|x - x'|,\\
&|g_2(x, \theta, z) - g_2(x, \theta', z)| \leq \psi_3(x)\|\theta - \theta'\|,
\end{aligned}$$

where $g_1 = \dot{\ell}, f', \ell''$ and $g_2 = \dot{\ell}, f, f', \ell', \ell''$.

The functions $\psi, \psi_p, p = 1, 2, 3$ may differ in each inequality. Note that we do not require differentiability of f with respect to θ, but only a Lipschitz continuity condition.

We shall estimate the parameter θ by solving the score equation $U_n(f, \theta) = o_P(n^{-1/2})$, where

$$U_n(f, \theta) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_0} \widetilde{W}_f(u, \theta, Z_i)\widetilde{M}_i(du, \theta)$$

and

$$\widetilde{M}_i(t, \theta) = N_i(t) - \int_0^t Y_i(u)\alpha(\Gamma_{n\theta}(u), \theta, Z_i)\Gamma_{n\theta}(du)$$

Here $\widetilde{W}_f$ is defined by substituting the estimate $\Gamma_{n\theta}$ into the score function $W_f(t, \theta, Z)$ of Proposition 2.1. A more explicit form of the score process is given in Section 5.

Let $\Sigma_0(f, \theta_0) = \Sigma_1(f, \theta_0) + \Sigma_2(f, \theta_0)$ be given by (2.10) and set

$$V(f, \theta_0) = \int_0^{\tau_0} \mathrm{cov}_0[f - \ell'\varphi_0, \dot{\ell} + \ell' D_0[\dot{\ell}]](u)EN(du),$$

where $\varphi_0 = \varphi^*$ is the solution to the Fredholm equation (2.6). Note that if $f(x, \theta, Z) = \dot{\ell}(x, \theta, Z)$, then

$$\begin{aligned}
V(f, \theta_0) &= \Sigma_1(f, \theta_0)\\
&+ \int \mathrm{cov}_0[f - \ell'\varphi_0, \ell'](u)(\varphi_0 + D_0[\dot{\ell}](u)EN(du)\\
&= \Sigma_1(f, \theta_0) + \Sigma_2(f, \theta_0)
\end{aligned}$$

Condition 4.3 (i) The matrix $\Sigma_0(f, \theta_0)$ is positive definite and the matrix $V(f, \theta_0)$ is non-singular.

(ii) The estimate $\varphi_{n\theta}$ of the solution to the Fredholm equation (2.6) satisfies $\limsup_n \sup\{\|\varphi_{n\theta}\|_v : \theta \in B(\theta_0, \varepsilon_n)\} = O_P(1)$ and $\sup\{\|\varphi_{n\theta} - \varphi_0\|_\infty : \theta \in B(\theta_0, \varepsilon_n)\} = o_P(1)$.

The form of the solution φ_0 to the equation (2.6) is given in Section 5.3. Therein we also verify the condition 4.3 (ii) for the sample counterpart of this equation based on an estimator $\Gamma_{n\theta}$ satisfying the conditions 4.1.

Proposition 4.1 Suppose that the conditions 2.1-2.3 and 4.1-4.3 hold. Then, with probability tending to 1, the score equation $U_n(f, \theta) = o_P(n^{-1/2})$ has a solution in $B(\theta_0, \varepsilon_n)$. In addition, $\sqrt{n}(\widehat{\theta} - \theta_0)$ converges in distribution to a multivariate normal variable $\mathcal{N}(0, \Sigma(f, \theta_0))$ with covariance matrix $\Sigma(f, \theta_0) = (V^{-1}\Sigma_0[V^{-1}]^T)(f, \theta_0)$.

In this proposition the asymptotic covariance matrix of the estimate does not depend on the estimate of the unknown transformation. In addition, if we choose $f = \dot{\ell}$, then proposition 2.1 entails $\Sigma(f, \theta_0) = \Sigma_0(f, \theta_0)$.

The second version of this proposition, assumes that a preliminary $\sqrt{n}$- consistent estimate $\widehat{\theta}^{(0)}$ of θ_0 is available. Define

$$\widehat{\theta} = \widehat{\theta}^{(0)} + V_n(f, \widehat{\theta}^{(0)})^{-1} U_n(f, \widehat{\theta}^{(0)})$$

Here $V_n(f, \widehat{\theta}^{(0)})$ is the plug-in estimate of the matrix $V_n(f, \theta_0)$. Section 5.2 gives the explicit form of this matrix.

Proposition 4.2 Suppose that the conditions 2.1-2.3 and 4.1-4.3 hold. Then $\sqrt{n}(\widehat{\theta} - \theta_0)$ converges in distribution to $\mathcal{N}(0, \Sigma(f, \theta_0))$.

Examples of simple $\sqrt{n}$ consistent estimators of the parameter θ were given in [5, 6, 7, 9, 12, 23].

5 Proof of Proposition 3.1

5.1 An auxiliary lemma

The proof of Proposition 3.1 is based on the following modification of Theorem 2 in Bickel *et al.* [4, p.518].

Lemma 5.1 Let $B(\theta_0, \varepsilon_n) = \{\theta : |\theta - \theta_0| \leq \varepsilon_n)$ be a ball of radius $\varepsilon_n, \varepsilon_n \to 0$, $\sqrt{n}\varepsilon_n \to \infty$. Suppose that

(i) $\sqrt{n}U_n(\theta_0) \Rightarrow \mathcal{N}(0, \Sigma_0(\theta_0))$.

(ii) $V_{2n}(\theta_0) \to_P V(\theta_0)$.

(iii) The matrices $\Sigma_0(\theta_0)$ and $V(\theta_0)$ are nonsingular.

(iv) $U_n(\theta) - U_n(\theta_0) = (\theta - \theta_0)^T V_n(\theta_0) + \mathrm{rem}(\theta)$, where

$$\sup\left\{\frac{\sqrt{n}|\mathrm{rem}(\theta) - \mathrm{rem}(\theta_0)|}{1 + \sqrt{n}|\theta - \theta_0|} : \theta \in B(\theta_0, \varepsilon_n)\right\} \to_P 0.$$

If the assumptions (i)-(iv) are satisfied then with probability tending to 1, the score equation $U_n(\theta) = o_P(n^{-1/2})$ has a solution $\widehat{\theta}$ in $B(\theta_0, \varepsilon_n)$ and

$$\sqrt{n}(\widehat{\theta} - \theta_0) \Rightarrow N(0, [V^{-1}\Sigma_0(V^T)^{-1}](\theta_0))$$

Proof . Let $\overline{U}_n(\theta) = U_n(\theta) - \mathrm{rem}(\theta)$. We have $\overline{U}_n(\theta_0) = U_n(\theta_0)$ because $\mathrm{rem}(\theta_0) = 0$. Set

$$\begin{aligned} a_n &= \|I - V^{-1}(\theta_0)V_n(\theta_0)\| = o_P(1), \\ A_n &= V_2^{-1}(\theta_0)\overline{U}_n(\theta_0) = O_P(n^{-1/2}). \end{aligned}$$

Finally, define $h_n(\theta) = \theta - V^{-1}(\theta_0)\overline{U}_n(\theta)$, and put $\theta_n^{(0)} = \theta_0$ and $\theta_n^{(m)} = h_n(\theta_n^{(m-1)})$ for $m \geq 1$. The condition (iv) implies that for $m \geq 1$ we have

$$\theta_n^{(m)} - \theta_n^{(0)} = [I - (V^{-1}V_n)(\theta_0)](\theta_n^{(m-1)} - \theta_n^{(0)}) - V^{-1}(\theta_0)\overline{U}_n(\theta_0) \tag{5.1}$$
$$h(\theta_n^{(m)}) - h(\theta_n^{(m-1)}) = [I - (V^{-1}V_n)(\theta_0)](\theta_n^{(m)} - \theta_n^{(m-1)}) \tag{5.2}$$

Similarly to Bickel et al ([4, p.518]), (5.1) - (5.2) implies that the mapping h_n is a contraction on the ball $B_n = \{\theta : |\theta - \theta_0| \le A_n/(1-a_n)\}$. With probability tending to 1, $B_n \subset B(\theta_0, \varepsilon_n)$, because $A_n = O_P(n^{-1/2}), a_n = o_P(1)$ and $\sqrt{n}\varepsilon_n \uparrow \infty$. It follows that with probability tending to 1, the equation $V^{-1}(\theta_0)\overline{U}_n(\theta) = 0$ has a unique root $\widehat{\theta}$ in B_n satisfying $\widehat{\theta} - \theta_0 = V_2(\theta_0)^{-1}\overline{U}_n(\theta_0) = O_P(n^{-1/2})$. We also have $U_n(\widehat{\theta}) = \overline{U}(\widehat{\theta}) + \mathrm{rem}(\widehat{\theta}) = o_P(|\widehat{\theta} - \theta_0| + n^{-1/2}) = o_P(O_P(n^{-1/2}) + n^{-1/2}) = o_P(n^{-1/2})$. $\square$

5.2 Proof of Proposition 3.1

Define

$$\begin{aligned}
\widehat{s}[1](t, \Gamma_{n\theta}, \theta) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t)\alpha(\Gamma_{n\theta}(t), \theta, Z_i), \\
\widehat{s}[f](t, \Gamma_{n\theta}, \theta) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t) f(\Gamma_{n\theta}(t), \theta, Z_i)\alpha(\Gamma_{n\theta}(t), \theta, Z_i), \\
\widehat{e}[f](t, \Gamma_{n\theta}, \theta) &= \frac{\widehat{s}[f]}{\widehat{s}[1]}(t, \Gamma_{n\theta}, \theta).
\end{aligned}$$

Similarly to section 2, we put

$$\begin{aligned}
\widehat{\mathrm{cov}}[f_1, f_2](t, \Gamma_{n\theta}, \theta) &= (\widehat{e}[f_1 f_2^T] - \widehat{e}[f_1]\widehat{e}[f_2^T])(t, \Gamma_{n\theta}, \theta), \\
\widehat{\mathrm{var}}[f](t, \Gamma_{n\theta}, \theta) &= \widehat{\mathrm{cov}}[f, f](t, \Gamma_{n\theta}, \theta).
\end{aligned}$$

Let $N_{.}(t) = n^{-1}\Sigma_{i=1}^{n} N_i(t)$ and set

$$\begin{aligned}
\mathcal{P}_{n\theta}(s, t) &= \exp - \int_s^t \widehat{s}[\ell'](u, \Gamma_{n\theta}, \theta) C_{n\theta}(du), \\
C_{n\theta}(t) &= \int_0^t \widehat{s}[1](u, \Gamma_{n\theta}, \theta)^{-2} N_{.}(du) \qquad (5.3) \\
\widehat{\rho}[f, \varphi_{n\theta}](t, \Gamma_{n\theta}, \theta) &= \widehat{\mathrm{cov}}[f, \ell'](t, \Gamma_{n\theta}, \theta) - \widehat{\mathrm{var}}[\ell'](t, \Gamma_{n\theta}, \theta)\varphi_{n\theta}(t),
\end{aligned}$$

where $\varphi_{n\theta}$ is an estimate of the solution to the Fredholm equation (2.6).

The score process for estimation of the parameter θ is given by

$$\widetilde{U}_n(f, \theta) = \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau_0} \widetilde{W}_f(t, \theta, Z_i)\widetilde{M}_i(dt, \theta),$$

where

$$\widetilde{M}_i(t,\theta) = N_i(t) - \int_0^t Y_i(u)\alpha(\Gamma_{n\theta}(u),\theta,Z_i)\Gamma_{n\theta}(du)$$

and

$$\begin{aligned}
\widetilde{W}_f(t,\theta,Z_i) &= b_{1i}(t,\Gamma_{n\theta},\theta) - b_{2i}(t,\Gamma_{n\theta},\theta)\varphi_{n\theta}(t),\\
&- [\widehat{s}[1](t,\Gamma_{n\theta},\theta)]^{-1}\int_t^{\tau_0}\mathcal{P}_{n\theta}(t,u)\widehat{\rho}[f,\varphi_{n\theta}](u,\Gamma_{n\theta},\theta)N_{\cdot}(du),\\
b_{1i}(t,\Gamma_{n\theta},\theta) &= f(\Gamma_{n\theta}(t),\theta,Z_i) - \widehat{e}[f](t,\Gamma_{n\theta},\theta),\\
b_{2i}(t,\Gamma_{n\theta},\theta) &= \ell'(\Gamma_{n\theta}(t),\theta,Z_i) - \widehat{e}[\ell'](t,\Gamma_{n\theta},\theta).
\end{aligned}$$

The form of the score process simplifies if we introduce

$$\widehat{\Gamma}_{n\theta}(t) = \int_0^t \frac{N_{\cdot}(du)}{\widehat{s}[1](u,\Gamma_{n\theta},\theta)}.$$

We have

$$\begin{aligned}
U_n(f,\theta) &= \frac{1}{n}\sum_{i=1}^n\int_0^{\tau_0}\left[b_{1i}(t,\Gamma_{n\theta},\theta) - b_{2i}(t,\Gamma_{n\theta},\theta)\varphi_{n\theta}(t)\right]N_i(dt)\\
&- \int_0^{\tau_0}\left[\int_t^{\tau_0}\mathcal{P}_{n\theta}(t,u)\widehat{\rho}[f,\varphi_{n\theta}](u,\Gamma_{n\theta},\theta)N_{\cdot}(du)\right][\widehat{\Gamma}_{n\theta}-\Gamma_{n\theta}](dt).
\end{aligned}$$

Set $M_{\cdot}(t) = n^{-1}\Sigma_{i=1}^n M_i(t)$. Then

$$\begin{aligned}
[\widehat{\Gamma}_{n0}-\Gamma_{n0}](t) &= \int_0^t \frac{M_{\cdot}(du)}{s[1](u,\Gamma_0(u),\theta_0)} - [\Gamma_{n0}-\Gamma_0](t)\\
&- \int_0^t [\Gamma_{n0}-\Gamma_0](u)e_0[\ell'](u)\Gamma_0(du) + o_p(n^{-1/2})\ . \qquad (5.4)
\end{aligned}$$

The score process $U_n(f,\theta_0)$ can be represented as a sum $U_n(f,\theta_0) = \sum_{j=1}^4 U_{nj}(f,\theta_0)$, where

$$\begin{aligned}
U_{n1}(f,\theta_0) &= \frac{1}{n}\sum_{i=1}^n\int_0^{\tau_0}[b_{1i}(t,\Gamma_0,\theta_0) - b_{2i}(t,\Gamma_0,\theta_0)\varphi_0(t)]N_i(dt),\\
U_{n2}(f,\theta_0) &= -\int_0^{\tau_0}[\widehat{\Gamma}_{n0}-\Gamma_{n0}](du)\int_u^{\tau}\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](t)EN_{\cdot}(dt) + o_P(n^{-1/2}),\\
U_{n3}(f,\theta_0) &= -\int_0^{\tau_0}[\Gamma_{n0}-\Gamma_0](t)\rho_0[f,\varphi_0](t)EN(dt) + o_P(n^{-1/2}),\\
U_{n4}(f,\theta_0) &= \int_0^{\tau_0}[\varphi_{n0}-\varphi_0](t)\frac{1}{n}\sum_{i=1}^n b_{2i}(\Gamma_0(t),\theta_0,t)N_i(dt) = o_P(n^{-1/2})\ .
\end{aligned}$$

By central limit theorem, we have $\sqrt{n}U_{n1}(f,\theta_0) \Rightarrow N(0,\Sigma_1(f,\theta_0))$. The matrix $\Sigma_1(f,\theta_0)$ is defined in Section 2. Further,

$$
\begin{aligned}
& U_{n2}(f,\theta_0) + U_{n3}(f,\theta_0) = -\int_0^{\tau_0} \left[\int_0^t [\widehat{\Gamma}_{n0} - \Gamma_{n0}](du)\mathcal{P}_0(u,t)\right] \rho_0[f,\varphi_0](t) EN(dt) \\
- & \int_0^{\tau_0} [\Gamma_{n0} - \Gamma_0](t)\rho_0[f,\varphi_0](t)EN(dt) + o_P(n^{-1/2}) = \\
= & \int_0^{\tau_0} \left[\int_0^t [\widehat{\Gamma}_{n0} - \Gamma_{n0}](du) \int_u^t e_0[\ell'](s)\Gamma_0(ds)\mathcal{P}_0(s,t)\right] \rho_0[f,\varphi_0](t)EN(dt) \\
- & \int_0^{\tau_0} \left[\int_0^t [\widehat{\Gamma}_{n0} - \Gamma_{n0}](du)\right] \rho_0[f,\varphi_0](t)EN(dt) \\
- & \int_0^{\tau_0} [\Gamma_{n0} - \Gamma_0](t)\rho_0[f,\varphi_0](t)EN(dt) + o_P(n^{-1/2}) \\
= & \int_0^{\tau_0} \left[\int_0^t [\widehat{\Gamma}_{n0} - \Gamma_{n0}](s)e_0[\ell'](s)\Gamma_0(ds)\mathcal{P}_0(s,t)\right] \rho_0[f,\varphi_0](t)EN(dt) \\
- & \int_0^{\tau_0} [\widehat{\Gamma}_{n0} - \Gamma_{n0}](t)\rho_0[f,\varphi_0](t)EN(dt) \\
- & \int_0^{\tau_0} [\Gamma_{n0} - \Gamma_0](t)\rho_0[f,\varphi_0](t)EN(dt) + o_P(n^{-1/2}) \; .
\end{aligned}
$$

Next substitution of (5.3) yields

$$
\begin{aligned}
& U_{n2}(f,\theta_0) + U_{n3}(f,\theta_0) = \\
- & \int_0^{\tau_0} \left[\int_0^t \left(\int_0^s [\Gamma_{n0} - \Gamma_0](v)e_0[\ell'](v)EN(dv)\right) e_0[\ell'](s)\Gamma_0(ds)\mathcal{P}_0(s,t)\right] \rho_0[f,\varphi_0](t)EN(dt) \\
- & \int_0^{\tau_0} \left[\int_0^t [\Gamma_{n0} - \Gamma_0](s)e_0[\ell'](s)\Gamma_0(ds)\mathcal{P}_0(s,t)\right] \rho_0[f,\varphi_0](t)EN(dt) \\
+ & \int_0^{\tau_0} \left[\int_0^t \left(\int_0^s \frac{dM_.}{s_0[1]}\right) e_0[\ell'](s)\Gamma_0(ds)\mathcal{P}_0(s,t)\right] \rho_0[f,\varphi_0](t)EN(dt) \\
+ & \int_0^{\tau_0} \left[\int_0^t [\Gamma_{n0} - \Gamma_0](u)e_0[\ell']EN(du)\right] \rho_0[f,\varphi_0](t)EN(dt) \\
- & \int_0^{\tau_0} \left(\int_0^t \frac{dM_.}{s_0[1]}\right) \rho_0[f,\varphi_0](t)EN(dt) + o_P(n^{-1/2}) \; .
\end{aligned}
$$

Using

$$
\mathcal{P}_0(u,t) - 1 = \int_u^t s_0[\ell'](v)C_0(dv)\mathcal{P}_0(v,t) = \int_u^t e_0[\ell'](v)\Gamma_0(dv)\mathcal{P}_0(v,t) \tag{5.5}
$$

and Fubini theorem, it is easy to see that the first, the second and the fourth term of this expansion sum to 0. The sum of the remaining terms is

$$U_{n2}(f,\theta_0)+U_{3n}(f,\theta_0)=-\int_0^{\tau_0}\frac{M_\cdot(du)}{s_0[1](u)}\left[\int_u^{\tau_0}\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](t)EN(dt)\right]+o_P(n^{-1/2})\ .$$

We have $\sqrt{n}[U_{n2}+U_{n3}](f,\theta_0)\Rightarrow N(0,\Sigma_2(f,\theta_0))$, and the matrix $\Sigma_2(f,\theta_0)$ is defined in Section 2. It is also easy to verify that $\sqrt{n}[U_{n2}+U_{n3}](f,\theta_0)$ and $\sqrt{n}U_{n1}(f,\theta_0)$ are asymptotically uncorrelated. Therefore $\sqrt{n}U_n(f,\theta_0)\Rightarrow\mathcal{N}(0,\Sigma_0(f,\theta_0))$, $\Sigma_0=\Sigma_1+\Sigma_2$.

We consider now the expansion of the score process $U_n(f,\theta)-U_n(f,\theta_0)$ for $\theta\in B(\theta_0,\varepsilon_n)$. Set $\widehat{W}(\theta)=\widehat{\Gamma}_{n\theta}-\Gamma_{n\theta}-\widehat{\Gamma}_{n0}+\Gamma_{n0}$. Then

$$\begin{aligned}\widehat{W}(\theta)(t) &= -(\theta-\theta')\int_0^t e_0[\dot{\ell}](u)\Gamma_0(du)-\int_0^t[\Gamma_{n\theta}-\Gamma_{n0}](u)e_0[\ell'](u)\Gamma_0(du)\\ &- [\Gamma_{n\theta}-\Gamma_{n0}](t)+\mathrm{Rem}(\theta)(t),\end{aligned}\tag{5.6}$$

where the remainder term satisfies

$$\sup\{\frac{\sqrt{n}|\mathrm{Rem}(\theta)(t)-\mathrm{Rem}(\theta_0)(t)|}{\sqrt{n}|\theta-\theta_0|+1}:\theta\neq\theta_0,\theta\in B(\theta_0,\varepsilon_n),t\le\tau_0\}=o_P(1)$$

For $\theta\in B(\theta_0,\varepsilon_n)$, we also have $\|\Gamma_{n\theta}-\Gamma_{n0}\|_\infty=o_P(|\theta-\theta_0|+n^{-1/2})$.

Define

$$\begin{aligned}\widetilde{e}[f](u,\Gamma_0,\theta,\theta_0) &= \frac{\Sigma_{i=1}^n Y_i(u)f(\Gamma_0(u),\theta,Z_i)\alpha(\Gamma_0(u),\theta_0,Z_i)}{n\widehat{s}[1](u,\Gamma_0,\theta_0)},\\ \widetilde{e}[\ell'](u,\Gamma_0,\theta,\theta_0) &= \frac{\Sigma_{i=1}^n Y_i(u)\ell'(\Gamma_0(u),\theta,Z_i)\alpha(\Gamma_0(u),\theta_0,Z_i)}{n\widehat{s}[1](u,\Gamma_0,\theta_0)}\end{aligned}$$

ans let

$$\begin{aligned}I_{1n}(\theta) &= \frac{1}{n}\sum_{i=1}^n\int_0^{\tau_0}\Big[f(\Gamma_0(u),\theta,Z_i)-f(\Gamma_0(u),\theta_0,Z_i)\\ &\qquad -\widetilde{e}[f](u,\Gamma_0,\theta,\theta_0)+\widehat{e}[f](u,\Gamma_0,\theta_0)\Big]N_i(du),\\ I_{2n}(\theta) &= -\frac{1}{n}\sum_{i=1}^n\int_0^{\tau_0}\Big[\ell'(\Gamma_0(u),\theta,Z_i)-\ell'(\Gamma_0(u),\theta_0Z_i)\\ &\qquad -\widetilde{e}[\ell'](u,\Gamma_0,\theta,\theta_0)+\widehat{e}[\ell'](u,\Gamma_0,\theta_0)\Big]\varphi_0(u)N_i(du).\end{aligned}$$

The condition 4.2 implies that $I_{qn}(\theta), q = 1, 2$ is a mean zero square integrable martingale and $\mathrm{Var}[\sqrt{n} I_{qn}(\theta)] = O(1)(\theta - \theta_0)^2 = O(\varepsilon_n^2) = o(1)$. Using Hoeffding projection method, we can show that the right hand side of these expressions can be approximated by a sum of U-processes of degree $k, k \le 4$ over Euclidean classes of functions with square integrable envelopes $H_{nk}, k \le 4$. Under the assumption of the transformation model, the L_2-norm of EH_{nk} is of order $O(\varepsilon_n)$. Application of maximal inequalities for U-processes indexed by Euclidean classes of functions [14] shows that $I_{qn}(\theta) = o_P(n^{-1/2})$, uniformly in $\theta \in B(\theta_0, \varepsilon_n)$. The details are similar to [12, 13], so we omit the proof.

The score process satisfies $U_n(f, \theta) - U_n(f, \theta_0) = \Sigma_{j=1}^7 I_{jn}(\theta)$, where $I_{1n}(\theta)$ and $I_{2n}(\theta)$ are defined as above and

$$\begin{aligned}
I_{3n}(\theta) &= \frac{1}{n}\sum_{i=1}^n \int_0^{\tau_0} [f(\Gamma_{n\theta}(u), \theta, Z_i) - f(\Gamma_{n0}(u), \theta_0, Z_i) \\
&\qquad -\widehat{e}[f](u, \Gamma_{n\theta}, \theta) + \widehat{e}[f](u, \Gamma_{n0}, \theta_0)] N_i(du) - I_{1n}(\theta) \\
&= \frac{1}{n}\sum_{i=1}^n \int_0^{\tau_0} [\Gamma_{n\theta} - \Gamma_{n0}](u)[f'(\Gamma_0(u), \theta_0, Z_i) - \widehat{e}[f'](u, \Gamma_0, \theta_0)] N_i(du) \\
&- (\theta - \theta_0) \int_0^{\tau_0} \widehat{\mathrm{cov}}[f, \dot\ell](u, \Gamma_0, \theta_0) N_{\cdot}(du) \\
&- \int_0^{\tau_0} [\Gamma_{n\theta} - \Gamma_{n0}](u) \widehat{\mathrm{cov}}[f, \ell'](u, \Gamma_0, \theta_0) N_{\cdot}(du) + o_P(|\theta - \theta_0| + \|\Gamma_{n\theta} - \Gamma_{n0}\|_\infty + n^{-1/2}) \\
&= -(\theta - \theta_0) \int_0^{\tau_0} \mathrm{cov}_0[f, \dot\ell](u) N_{\cdot}(du) - \int_0^{\tau_0} [\Gamma_{n\theta} - \Gamma_{n0}] \mathrm{cov}_0[f, \ell'](u) N_{\cdot}(du) \\
&+ o_p(|\theta - \theta_0| + n^{-1/2}) \\
I_{4n}(\theta) &= -\frac{1}{n}\sum_{i=1}^n \int_0^{\tau_0} [\ell'(\Gamma_{n\theta}(u), \theta, Z_i) - \ell'(\Gamma_{n0}(u), \theta_0, Z_i) \\
&\qquad -\widehat{e}[\ell'](u, \Gamma_{n\theta}, \theta) + \widehat{e}[\ell'](u, \Gamma_{n0}, \theta_0)] \varphi_0(u) N_i(du) - I_{2n}(\theta) \\
&= -\frac{1}{n}\sum_{i=1}^n \int_0^{\tau_0} [\Gamma_{n\theta} - \Gamma_{n0}](u)[\ell''(\Gamma_0(u), \theta_0, Z_i) - \widehat{e}[\ell''](u, \Gamma_0, \theta_0)] \varphi_0(u) N_i(du) \\
&+ (\theta - \theta_0) \int_0^{\tau_0} \widehat{\mathrm{cov}}[\ell', \dot\ell](u, \Gamma_0, \theta_0) \varphi_0(u) N_{\cdot}(du) \\
&+ \int_0^{\tau_0} [\Gamma_{n\theta} - \Gamma_{n0}](u) \widehat{\mathrm{var}}[l'](u, \Gamma_0, \theta_0) \varphi_0(u) N_{\cdot}(du) + o_P(|\theta - \theta_0| + \|\Gamma_{n\theta} - \Gamma_{n0}\|_\infty + n^{-1/2}) \\
&= (\theta - \theta_0) \int_0^{\tau_0} \mathrm{cov}_0[\ell', \dot l](u) \varphi_0(u) N_{\cdot}(du) + \int_0^{\tau_0} [\Gamma_{n\theta} - \Gamma_{n0}] \mathrm{var}_0[\ell'](u) \varphi_0(u) N_{\cdot}(du) \\
&+ o_p(|\theta - \theta_0| + n^{-1/2}).
\end{aligned}$$

Combining,

$$
\begin{aligned}
\sum_{j=1}^{4} I_{jn}(\theta) &= -(\theta-\theta_0)\int_0^{\tau_0} \mathrm{cov}_0[f-\ell'\varphi_0,\dot\ell](u)N_{\cdot}(du) \\
&- \int_0^{\tau_0}[\Gamma_{n\theta}-\Gamma_{n0}](u)\rho_0[f,\varphi_0](u)N_{\cdot}(du)+o_P(|\theta-\theta_0|+n^{-1/2}).
\end{aligned}
$$

The remaining three terms of the expansion are given by

$$
\begin{aligned}
I_{n5}(f,\theta) &= -\int_0^{\tau_0}\widehat{W}(\theta)(du)\int_u^{\tau_0}\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](u)N_{\cdot}(du) \\
&= (\theta-\theta_0)\int_0^{\tau_0} e_0[\dot\ell]\Gamma_0(du)\int_u^{\tau_0}\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](u)N_{\cdot}(du) \\
&+ \int_0^{\tau_0}[\Gamma_{n\theta}-\Gamma_{n0}](du)\int_u^{\tau_0}\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](t)N(dt) \\
&+ \int_0^{\tau_0}[\Gamma_{n\theta}-\Gamma_{n0}](u)e_0[\ell'](u)\Gamma_0(du)\int_u^{\tau_0}\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](u)N(du) \\
&+ o_P(|\theta-\theta_0|+\|\Gamma_{n\theta}-\Gamma_{n0}\|) \\
&= -(\theta-\theta_0)\int_0^{\tau_0} D_0[\dot\ell](u)\rho_0[f,\varphi_0](u)N_{\cdot}(du) \\
&+ \int_0^{\tau_0}[\Gamma_{n\theta}-\Gamma_{n0}](u)\rho_0[f,\varphi_0](f)N_{\cdot}(du)+o_P(|\theta-\theta_0|+n^{-1/2}), \\
I_{n6}(f,\theta) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_0}[b_{2i}(u,\Gamma_{n\theta}(u),\theta)-b_{2i}(u,\Gamma_{n0},\theta_0)][\varphi_{n\theta}-\varphi_0]N.(du) \\
&= o_P(|\theta-\theta_0|+\|\Gamma_{n\theta}-\Gamma_{n0}\|)=o_p(|\theta-\theta_0|+n^{-1/2}), \\
I_{n7}(f,\theta) &= -\int_0^{\tau_0}\widehat{W}(\theta)(du) \\
&\times \int_u^{\tau_0}[\mathcal{P}_{n\theta}(u,t)\widehat{\rho}[f,\varphi_{n\theta}](t,\Gamma_{n\theta},\theta)-\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](t)]N_{\cdot}(dt) \\
&- \int_0^{\tau_0}[\widehat{\Gamma}_{n0}-\Gamma_{n0}](du) \\
&\times \int_u^{\tau_0}[\mathcal{P}_{n\theta}(u,t)\widehat{\rho}[f,\varphi_{n\theta}](t,\Gamma_{n\theta},\theta)-\mathcal{P}_{n0}(u,t)\widehat{\rho}[f,\varphi_0](t,\Gamma_{n0},\theta_0)]N_{\cdot}(dt) \\
&+ \int_0^{\tau_0}[\widehat{\Gamma}_{n0}-\Gamma_{n0}](du) \\
&\times \int_u^{\tau_0}[\mathcal{P}_{n0}(u,t)\widehat{\rho}[f,\varphi_{n0}](t,\Gamma_{n0},\theta_0)-\mathcal{P}_0(u,t)\rho_0[f,\varphi_0](t)]N_{\cdot}(dt) \\
&= o_P(|\theta-\theta_0|+n^{-1/2})\ .
\end{aligned}
$$

Using (5.4) and (5.5), we find that

$$\begin{aligned} U_n(f,\theta) - U_n(f,\theta_0) &= \Sigma_{j=1}^{7} I_{jn}(f,\theta) = -(\theta-\theta_0)V_n(f,\theta_0) + o_P(|\theta-\theta_0| + n^{-1/2}), \\ V_n(f,\theta_0) &= \int_0^\tau \mathrm{cov}_0[f - \ell'\varphi_0, \dot{\ell} + \ell' D_0[\dot{\ell}]](u) N_{\cdot}(du). \end{aligned}$$

The matrix $V_n(f,\theta_0)$ converges in probability to the matrix $V(\theta_0)$ defined in Section 4. so that Lemma 5.1 completes the proof. $\square$

The proof of Proposition 4.2 follows from a similar expansion. The matrix $\widehat{V}_n(f,\theta^{(0)})$ can be defined by plugging-in the sample counterpart of the covariance operator in the last display.

5.3 Verification of the condition 4.3

We have shown in [12] that the equation (2.10) simplifies if we multiply both sides of it by $\mathcal{P}(0,t)^{-1}$. Let $\widetilde{\psi}(t) = \mathcal{P}(0,t)^{-1}\psi(t)$, $\widetilde{D}[f](t) = \mathcal{P}(0,t)^{-1}D[f](t)$ and $\widehat{\rho}[f,-D[f]](t) = \mathcal{P}(0,t)\rho[f,-D[f]](t)$ Set

$$c(t) = \int_0^t \mathcal{P}(0,u)^{-2} dC(u), \quad b(t) = \int_0^t \mathcal{P}(0,u)^2 B(du).$$

Multiplication of (2.7) by $\mathcal{P}(0,t)^{-1}$ yields

$$\widetilde{\psi}(t) + \int_0^\tau k(t,u)\widetilde{\psi}(u) b(du) = \int_0^\tau k(t,u)\widetilde{\rho}[f,-D[f]](u) EN(du) \;, \tag{5.7}$$

where the kernel k is given by $k(t,u) = c(t \wedge u)$. The square integrability condition 2.3 is equivalent to the assumption that $\kappa(\tau_0) = \int_0^{\tau_0} c(u) b(du)$ is finite. The solution to the equation (5.6) is given by

$$\widetilde{\psi}(t) = \int_0^\tau \widetilde{\Delta}(t,u)\widetilde{\rho}[f,-D[f](u) EN(du) \;, \tag{5.8}$$

where $\widetilde{\Delta}(t,u) = \widetilde{\Delta}(t,u,-1)$, and $\widetilde{\Delta}(t,u,\lambda)$ is the resolvent corresponding to the kernel k. The solution to the equation (2.6) is given by

$$\varphi(t) = -D[f](t) + \int_0^{\tau_0} \widetilde{\Delta}(t,u)\rho[f,-D[f]](u) EN(du)\mathcal{P}(0,u)\mathcal{P}(0,t) \tag{5.9}$$

From [12], $\widetilde{\Delta}(u,t)$ is given by

$$\widetilde{\Delta}(u,t) = \frac{\Psi_1(0, u \wedge t)\Psi_0(u \vee t, \tau_0)}{\Psi_0(0,\tau_0)}, \tag{5.10}$$

where for $s < t$, the interval functions $\Psi_0(s,t)$ and $\Psi_1(s,t)$ are defined as solutions to the Volterra equations

$$\begin{aligned}
\Psi_0(s,t) &= 1 + \int_{(s,t]} c((s,u])b(du)\Psi_0(u,t) = 1 + \int_{(s,t]} \Psi_0(s,u-)c(du)b([u,t]) \\
\Psi_1(s,t) &= c((s,t]) + \int_{(s,t]} c((s,u])b(du)\Psi_1(u,t) = c((s,t]) + \int_{(s,t]} \Psi_1(s,u)b(du)c((u,t]).
\end{aligned}$$

Define also

$$\begin{aligned}
\Psi_2(s,t) &= 1 + \int_{[s,t)} b([s,u))c(du)\Psi_2(u,t) = 1 + \int_{[s,t)} \Psi_2(s,u+)b(du)c((u,t)) \\
\Psi_3(s,t) &= b([s,t)) + \int_{[s,t)} b([s,u))c(du)\Psi_3(u,t) = b([s,t)) + \int_{[s,t)} \Psi_3(s,u)c(du)b([u,t))
\end{aligned}$$

Then

$$\begin{aligned}
\Psi_0(s,t) &= 1 + \int_{(s,t]} \Psi_1(s,u)b(du) = 1 + \int_{(s,t]} c(du)\Psi_3(u,t+)\,, \\
\Psi_1(s,t) &= \int_{(s,t]} \Psi_0(s,u-)c(du) = \int_{(s,t]} c(du)\Psi_2(u,t+)\,, \\
\Psi_2(s,t) &= 1 + \int_{[s,t)} b(du)\Psi_1(u,t-) = 1 + \int_{[s,t)} \Psi_3(s,u)c(du)\,, \\
\Psi_3(s,t) &= \int_{[s,t)} \Psi_2(s,u)b(du) = \int_{[s,t)} b(du)\Psi_0(u,t-)\,.
\end{aligned}$$

If τ_0 is an atom of the survival function $P(X > t)$, then $\Psi_j, j = 0,1,2,3$ form bounded monotone increasing interval functions. In particular, $\Psi_0(s,t) \leq \exp\kappa(\tau_0)$ and $\Psi_1(s,t) \leq \Psi_0(s,t)[c(t) - c(s)]$. If τ_0 is a continuity point of the survival function $P(X > t)$ and $\kappa(\tau_0) < \infty$, then $\Psi_0(s,t) \leq \exp\kappa(\tau_0)$ for any $0 < s < t \leq \tau_0$, while the remaining functions are locally bounded [12].

Next suppose that the transformation model holds with $(\theta,\Gamma) = (\theta_0,\Gamma_0)$. We assume that the estimate $\Gamma_{n\theta}$ satisfies the conditions 4.1 and show the natural plug-in

estimator $\varphi_{n\theta}$ of (5.9) satisfies the conditions 4.3. Define

$$\begin{aligned} c_{n\theta}(t) &= \int_0^t \mathcal{P}_{n\theta}(0,u)^{-2} C_{n\theta}(du) \\ b_{n\theta}(t) &= \int_0^t \mathcal{P}_{n\theta}(0,u)^{2} \widehat{\mathrm{var}}[\ell'](u,\Gamma_{n\theta},\theta) N_{\cdot}(du), \end{aligned}$$

where $C_{n\theta}$ and $\mathcal{P}_{n\theta}$ are defined as in Section 5.2. Then the sample analogue of the equation (5.9) reduces to a system of linear equations which can be solved by inverting a bandsymmetric tridiagonal matrix [11, 12].

Denote by $\Psi_{n\theta,j}$ the sample counterparts of the interval functions Ψ_j, $j = 1,2,3,4$. Using Fubini theorem we can show that

$$\begin{aligned} [\Psi_{n\theta,0} - \Psi_0](s,t) &= \int_{(s,t]} \Psi_{n\theta,0}(s,u-) c_{n\theta}(du) [b_{n\theta} - b]([u,t]) \\ &+ \int_{(s,t]} [c_{n\theta} - c)]((s,u]) b(du) \Psi_0(u,t) \\ &+ \int_{s<u_1<u_2\le t} \Psi_{n\theta,0}(s,u_1-) c_{n\theta}(du_1) [b_{n\theta} - b]([u_1,u_2)) c(du_2) \Psi_3(u_2,t+) \\ &+ \int_{s<u_1<u_2\le t} \Psi_{n\theta 1}(s,u_1) b_{n\theta}(du_1) [c_{n\theta} - c]((u_1,u_2]) b(du_2) \Psi_0(u_2,t) \end{aligned}$$

and

$$\begin{aligned} [\Psi_{n\theta,1} - \Psi_1](s,t) &= [c_{n\theta} - c]((s,t]) \\ &+ \int_{(s,t]} \Psi_{n\theta,1}(s,u) b_{n\theta}(du) [c_{n\theta} - c]((u,t]) \\ &+ \int_{(s,t]} [\Psi_{n\theta,0} - \Psi_0](s,u-) c(du) \end{aligned}$$

Under assumptions of the condition 4.1, we have $c_n = \sup\{|c_{n\theta} - c|(t) : \theta \in B(\theta_0,\varepsilon_n), t \le \tau_0\} \to_P 0$ and $b_n = \sup\{|b_{n\theta} - b|(t\pm) : \theta \in B(\theta_0,\varepsilon_n), t \le \tau_0\} \to_P 0$. In addition, for $q = 0,1$, we have

$$\begin{aligned} &\limsup_n \sup\{\Psi_{n\theta,q}(0,\tau) : \theta \in B(\theta_0,\varepsilon_n)\} \\ &\quad \le [c(\tau_0) - c(0)]^q \exp[\kappa(\tau_0)](1 + o_p(1)) = O_p(1) \end{aligned}$$

Hence

$$\widehat{r} \;=\; \sup\{|\Psi_{n\theta,0} - \Psi_0|(s,t) : 0 < s < t \le \tau, \theta \in B(\theta_0,\varepsilon_n)\}$$

$$\begin{aligned}
&\leq 2b_n \Psi_{n\theta,0}(0,\tau_0)\Psi_0(0,\tau_0) + 2c_n \Psi_{n\theta,1}(0,\tau)\Psi_3(0,\tau) \\
&\quad \sup\{|\Psi_{n\theta,1} - \Psi_1|(s,t) : 0 < s < t \leq \tau, \theta \in B(\theta_0, \varepsilon_n)\} \\
&\leq 2c_n \Psi_{n\theta,0}(0,\tau_0) + \widehat{r}c(\tau_0)
\end{aligned}$$

and both terms converge in probability to 0.

Let $\widetilde{\Delta}_{n\theta}(s,t)$ be defined similarly to (5.10) The preceding calculations, imply that $\sup\{\widetilde{\Delta}_{n\theta}(s,t) : s,t \in [0,\tau_0], \theta \in B(\theta_0,\varepsilon_n)\} = O_P(1)$ and $\sup\{|\widetilde{\Delta}_{n\theta} - \widetilde{\Delta}|(s,t) : s,t \in [0,\tau_0], \theta \in B(\theta_0,\varepsilon_n)\} = o_P(1)$. Verification that the sample analogue $\varphi_{n\theta}$ of the equation (5.9) satisfies the conditions 4.3 can be completed using Gronwall's inequalities given in [12] and integration by parts.

References

[1] Bennett. S. (1983). Analysis of the survival data by the proportional odds model. *Statistics in Medicine* **2** 273–277.

[2] Bickel, P. J. (1986) Efficient testing in a class of transformation models. In *Proceedings of the 45th Session of the International Statistical Institute* 23.3-63-23.3-81. ISI, Amsterdam.

[3] Bickel, P. J. and Ritov, Y. (1995). Local asymptotic normality of ranks and covariates in transformation models. In *Festschrift for L. LeCam* (D. Pollard and G. Yang, eds) Springer.

[4] Bickel, P., Klaassen, C., Ritov, Y. and Wellner, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins Univ. Press.

[5] Bogdanovicius, V. and Nikulin, M. (1999). Generalized proportional hazards model based on modified partial likelihood. *Lifetime Data Analysis* **5**, 329-350.

[6] Bogdanovicius, M. Hafdi, M. A. and Nikulin, M. (2004). Analysis of survival data with cross-effects of survival functions. *Biostatistics* **5** 415-425.

[7] Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *J. Amer. Statist. Assoc.* **92** 227-235.

[8] Cox, D. R. (1972). Regression models in life tables. *J. Roy. Statist. Soc. Ser. B.* **34** 187–202.

[9] Cuzick, J. (1988) Rank regression. *Ann. Statist.* **16** 1369–1389.

[10] Dabrowska, D. M. and Doksum, K.A. (1988). Partial likelihood in transformation models. *Scand. J. Statist.* **15** 1-23.

[11] Dabrowska, D. M. (2005). Quantile regression in transformation models. *Sankhya* **67** 153-187.

[12] Dabrowska, D. M. (2006). Estimation in a class of semiparametric transformation models. In *Second Erich L. Lehmann Symposium–Optimality.* (J. Rojo, Ed.) Institute of Mathematical Statistics, Lecture Notes and Monograph Series, **49** 166-216.

[13] Dabrowska, D. M. (2006). Estimation in a semiparametric Marshall-Olkin model. (manuscript)

[14] de la Pena, V. and Giné, E. (1999). *Decoupling: From Dependence to Independence.* Springer Verlag.

[15] Klaassen, C. A. J. (1993). Efficient estimation in the Clayton-Cuzick model for survival data. Tech. Report, University of Amsterdam, Amsterdam, Holland.

[16] Kosorok, M. R. , Lee, B. L. and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *Ann. Statist.* **32** 1448-1449.

[17] Lehmann, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23-43.

[18] Murphy, S. A., Rossini, A. J. and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.* **92**, 968–976.

[19] Nan, B., Edmond, M. and Wellner, J. A. (2004). Information bounds for Cox regression with missing data. *Ann. Statist.* **32** 723-753.

[20] Sasieni, P. (1992). Information bounds for the conditional hazard ratio in a nested family of regression models. *J. Roy. Statist. Soc. Ser.* **B 54** 617-635.

[21] Scharfstein, D. O., Tsiatis, A. A. and Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds–rate class of regression models for right–censored time to event data. *Lifetime and Data Analysis* **4** 355–393.

[22] Slud, E. and Vonta, F. (2004) Consistency of the NMPL estimator in the right censored transformation model. *Scand. J. Statist.* **31** 21–43.

[23] Yang, S. and Prentice, R. (1999). Semiparametric inference in the proportional odds regression model. *J. Amer. Statist. Assoc.* **94** 125-136.